OsteoArthritis and Cartilage (2006) **14,** 830–838

doi:10.1016/j.joca.2006.02.007

Osteoarthritis and Cartilage

International Cartilage Repair Society

OARSI OSTEOARTHRITIS RESEARCH SOCIETY INTERNATIONAL

Identification and cross-species comparison of canine osteoarthritic gene regulatory *cis*-elements

S. S. Hannenhalli Ph.D.†[a], R. P. Middleton Ph.D.‡*[a], S. Levy Ph.D.§, B. Perroud Ph.D.‖, J. A. Holzwarth Ph.D.¶, K. McDonald Ph.D.†† and S. S. Hannah Ph.D.‡

† *Department of Genetics, University of Pennsylvania, Philadelphia, PA 19104, USA*
‡ *Nestlé Purina Research, St. Louis, MO 63164, USA*
§ *J. Craig Venter Institute, Rockville, MD 20850, USA*
‖ *Applied Biosystems, Foster City, CA 94404, USA*
¶ *Nestlé Research Center, 1000 Lausanne, Switzerland*
†† *Applied Biosystems, Foster City, CA 94404, USA*

Summary

Objective: To better understand transcription regulation of osteoarthritis (OA) by examining common promoter motifs in canine osteoarthritic genes, to identify other genes containing these motifs and to assess the conservation of these motifs between canine, human, mouse and rat.

Design: Differentially expressed transcripts in canine OA were mapped to the human genome. We thus identified 20 orthologous human transcripts representing 19 up-regulated genes and 62 orthologous transcripts representing 60 down-regulated genes. The 5 kbp upstream regions of these transcripts were used to identify binding sites and build promoter models based on those sites. The human genome was subsequently searched for other transcripts likely to be regulated by the same promoter models. Orthologous transcripts were then identified in canine, rat and mouse for determination of potential cross-species conservation of binding sites comprising the promoter model.

Results: Four promoter models containing 5–6 transcripts and 5–8 common transcription factor binding sites were developed. They include binding sites for AP-4, AP-2α and γ, and E2F. Several hundred other human genes were found to contain these promoter motifs. Furthermore these motifs were significantly over represented in the orthologous genes in canine, rat and mouse genomes.

Conclusions: We have developed and applied a computational methodology to identify common promoter elements implicated in OA and shared amongst four higher vertebrates. The transcription factors associated with these binding sites and other genes driven by these promoter motifs have been implicated in OA, chondrocyte development and with other biological factors involved in the disease.

Key words: Osteoarthritis, Chondrocyte, Gene expression, Promoter.

Introduction

Osteoarthritis (OA) affects 20% of the general canine population and 12% of the human population aged 25–74[1,2]. OA is a chronic degenerative disease characterized by the breakdown of articular cartilage. Inflammation is sometimes present, but not always. The degeneration of cartilage is a result of cartilage matrix dynamics being shifted in favor of catabolism over anabolism.

Molecular research of OA has included the identification of multiple changes at the protein or transcript level simultaneously. Key targets identified include matrix metalloproteinases (MMPs), specific tissue inhibitors of metalloproteinases (TIMPs), proinflammatory cytokines, insulin-like growth factors (IGFs) and other players known to be involved in cartilage metabolism, inflammatory pathways and bone formation. Groups of these are involved in common biological pathways and are likely to be driven by common regulatory mechanisms involving transcription factors and their associated DNA binding sites in gene regulatory regions. Thus the elucidation of significant motifs in these regulatory regions would give us a greater understanding of the mechanisms of disease etiology.

We had previously identified approximately 220 differentially expressed osteoarthritic genes using a canine OA microarray developed from differential display.[3] We describe here the development of promoter models by using *cis*-regulatory elements associated with the human orthologs of these genes. In addition, the human genome was searched for other genes likely to be driven by the same promoter elements in an attempt to identify other genes potentially involved in OA pathogenesis. The human genes thus identified contain sets of *cis*-regulatory elements that are also found in corresponding regions of the orthologous genes in the canine, mouse and rat genomes, suggesting the conservation of gene function in degenerative joint disease.

[a]The first two authors contributed equally.

*Address correspondence and reprint requests to: R. P. Middleton, Ph.D., Nestle Purina Research, Checkerboard Square, 2S St. Louis, MO 63164, USA. Tel: 1-314-982-2502; Fax: 1-314-982-5857; E-mail: rondo.middleton@rdmo.nestle.com

Received 8 September 2005; revision accepted 9 February 2006.

Materials and methods

PROMOTER MODEL DEVELOPMENT

The computational pipeline developed for the analysis is depicted in Fig. 1. Approximately 220 differentially expressed canine OA transcripts were mapped to human transcripts and the corresponding human transcripts were used to identify common transcription factor binding sites 5 kb upstream of the transcription start sites[4]. Multiple transcripts for a gene were used if the 5′ ends of the (alternative) transcripts were greater than 1500 bp apart. This ensured the inclusion of representative sets of transcripts while avoiding redundancy. Promoter model identification was performed separately for up and down-regulated genes. Twenty human transcripts derived from 19 genes that were up-regulated and 62 transcripts derived from 60 genes that were down-regulated were used. A graph theoretic approach for transcription model discovery was used as described[5]. Briefly, given the set of upstream regions and the set of potential binding sites in these regions, our approach is to identify large subsets of genes sharing a large set of common binding sites in their upstream regions. This can be done using algorithms for a well-studied problem of *clique-enumeration* from the computer science literature. The binding site annotation was based on known transcription factor binding specificities from TRANSFAC database[6] as well as human–mouse conservation as described[7]. Each of the four identified models included 5–6 transcripts containing 5–8 common transcription factor binding sites in their 5 kbp upstream regions (Tables I and II).

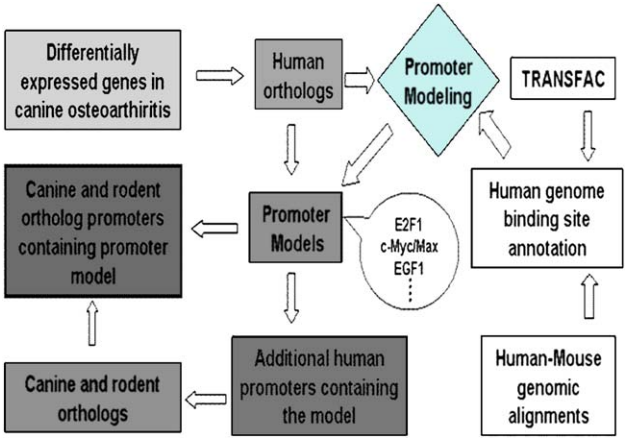

Fig. 1. The computational pipeline for transcriptional analysis of OA. We employed a previously described method for whole genome prediction of transcription factor binding sites[7]. Differentially expressed genes in canine OA are mapped to human transcripts. The corresponding 5 kb region upstream of the gene start for these human orthologs is used for promoter modeling as described previously[5]. The promoter models consist of sets of transcription factors which are then used to identify additional human genes which are likely to be under similar coordinated regulation and thus putatively involved in OA. In order to determine whether this set of putative, model identified, gene set is likely to be co-coordinately regulated they are used to identify orthologous genes in dog, mouse and rat. The upstream 5 kb regions of these orthologous canine and rodent genes are searched for the promoter model to evaluate the conservation of promoter model across species.

IDENTIFICATION OF OTHER GENES CONTAINING THE PROMOTER MODELS

After the transcriptional models were developed, the human genome was scanned for other transcripts likely to be regulated by the same models, as evidenced by the presence of potential binding sites in their 5 kb upstream regions. After the discovery of the initial set of such transcripts, an additional filtering step was employed. Only the transcripts with an average *cis*-similarity (averaged over the *cis*-similarities against the initial differentially expressed transcript) of at least 2 standard deviations above the mean were retained. *Cis*-similarity is a measure of commonality between the *cis*-elements in the upstream region of two genes. The mean and standard deviation of *cis*-similarity were based on a large set of unrelated transcript pairs. The retained transcripts for each model now represent the *output*.

DOG, MOUSE AND RAT ORTHOLOGS

For each of the four models, the output set of human transcripts was used to identify the canine orthologs using BLAST[8]. Orthologs were accepted when a sufficient length of match was obtained between the human transcript and dog genomic sequence (minimum 70% sequence identity for 80 nucleotides) and when the 5′ end of the human transcript aligned well. In particular, orthologs were discarded when more than 30 bp of the 5′ end of the human transcript was not aligned. The identification of the mouse and rat orthologs of human genes proceeded in a different manner than for canine due to the presence of independently annotated gene, transcript and protein set for those genomes (www.ensembl.org). Mutually best protein alignments were identified between the human–mouse and human–rat sequence sets by using BLASTP and identifying all proteins where greater than 50% of their length occurred in alignment blocks of 25 amino acids at greater than 65% sequence identity. We thus identified 14,326 human–mouse and 13,294 human–rat orthologs. These protein orthologs could be trivially associated to transcripts to identify the corresponding 5 kbp upstream regions for promoter analysis.

CONSERVATION OF PROMOTER MODELS IN DOG, MOUSE AND RAT

Five kilobase (or less when unavailable) upstream of the putative canine (similarly for mouse and rat) transcripts were searched for the transcriptional model (presence of potential binding sites) in these regions. Redundancy among the upstream regions was removed using the nonredundant database (nrdb) program (http://blast.wustl.edu/pub/nrdb/executables/). Unlike the search on human transcripts for the initial model construction, the conservation criterion of the transcription factor binding site hits was not used. Also there is inherent redundancy among the binding site profiles represented in the TRANSFAC database. For instance many binding site profiles correspond to same biological factor, e.g., Sp1 and many profiles corresponding to different factors are still very similar. Such a redundancy in the motifs representing a model would lead to wrong estimation of the enrichment. We have developed an approach to remove such redundant profiles. For example, down-regulated model 2 has eight factors – M00008 (Sp1), M00118 (cMyc/Max), M00189 (AP-2), M00243 (Egr-1), M00244 (NGFI-C), M00255 (GC_Box), M00324 (Muscle_initiator_sequence-20), and M00333 (ZF5). Four of these

Table I
The human orthologs of differentially regulated canine transcripts in the discovered promoter models. X denotes transcripts associated with each individual model

Gene symbol	Description	Up-regulated model 1	Down-regulated model 1	Down-regulated model 2	Down-regulated model 3
SMPD3	Sphingomyelin phosphodiesterase	X			
Q9BYE2_HUMAN	Protease, serine, 11 (IGF-binding)	X			
CHAD	Chondroadherin	X			
DGCR2	DiGeorge syndrome critical region gene 2	X			
SLC39A7	HLA class II region gene KE4 aka solute carrier family 39 (zinc transporter), member 7	X			
NFAT2	Nuclear factor of activated T-cells 5		X		X
SOX-9	SRY-box 9		X		X
HELZ	Helicase with zinc finger domain		X		
THBS1	Thrombospondin		X	X	X
VAV1	VAV proto-oncogene related		X	X	X
HK2	Hexokinase 2		X	X	X
ARID4A	RBP1			X	
ATXN1	Spinocerebellar ataxia 1			X	
PCSK5	Proprotein convertase subtilisin/kexin type 5			X	X

profiles are very similar to each other, three of which were removed, resulting in a set of five profiles in this model.

A promoter was considered ''hit'' by a model if every transcription factor in the model had a binding site in the promoter region. For each of the four models, the upstream sequences (of orthologous genes) were searched for the transcription factors comprising the corresponding model using the pwm_scan program with a *P*-value threshold of -9.21[7]. At this threshold we expect to see a hit once every 5000 bp (on either strand) on average i.e., the chance of a hit at any given position is 1/5000. Assuming positional independence of hits with positional hit probability $P = 1/5000$ and upstream region length L, the probability of observing at least one hit in the region (on either strand) is $P = 1 - (1 - 1/5000)^L$. If there are F profiles in the model, assuming independence among these profiles, the probability of observing a hit for each of the F profiles is P^F. If there are N orthologous genes with an average upstream

Table II
Transcription factors associated with the up- and down-regulated promoter models as inferred by the presence of their cognate binding sites in the promoters of transcripts in the model (Table I). X denotes transcription factors associated with each individual model. Note that each model contains a distinct set of transcription factors and those comprising the up-regulated model are quite different from those of the three down-regulated models

Transcription factor	Up-regulated model 1	Down-regulated model 1	Down-regulated model 2	Down-regulated model 3
Activating enhancer binding protein 4	X			
GATA binding protein 1	X			
GATA binding protein 2	X			
Muscle initiator sequence-19	X			
Muscle initiator sequence-20	X	X	X	X
Testis specific protein NYD-TSP1	X			
SP1		X	X	X
GC box		X	X	X
Zinc finger protein 161 homolog		X	X	X
Activating enhancer binding protein 2 alpha		X		X
Activating enhancer binding protein 2 gamma		X	X	X
E2F1		X		X
cMyc/Max			X	X
EGR 1			X	
EGR 4			X	

region length of L, we calculate the expected number of regions containing a hit for each of the F profiles as $E = N \times P^F$. Let the actual number of upstream regions hit by all F profiles be O, then the enrichment of the model in the particular species (dog, mouse or rat) is O/E. Table III shows the results of this enrichment analysis for dog, mouse and rat. Columns 2–4 show N, O, E, and O/E. After removing redundant positional weight matrices (PWMs) in the models as described earlier, $F = 5$ for up-regulated model 1 and $F = 6$, 5, and 6 for down-regulated models 1–3, respectively. $L = 3700$ bps (on average, based on sequence availability) for dog, and 5000 bps for mouse and rat.

GENE ONTOLOGY AND BIOLOGICAL ASSOCIATION ANALYSIS

Significant Gene Ontology biological processes represented by human genes containing the promoter models were determined by calculating a P-value which indicated which biological processes were over represented in the list of selected genes[9]. Only those processes where $P < 0.05$ and greater than 2 nodes were represented in a branch were retained (Table IV). Biological associations were determined for each set of human genes containing the promoter models using the software program Pathway Assist (Stratagene, USA). Associations (regulation, transport, binding, etc.) were assigned using natural language scans of Medline abstracts and databases such as DIP and BIND. Down-regulated promoter model 2 protein associations were developed by displaying only direct interactions (Fig. 2).

Results

PROMOTER MODELS

The discovered promoter models identified many transcription factors known to be associated with osteoarthritic genes but also included a few novel ones. One model was detected for up-regulated transcripts (Table II). However, three models for down-regulated transcripts had the same number of genes and transcription factors and hence all three of them were considered. The up-regulated model contains binding sites for transcription factors activating enhancer binding protein (AP) 4, GATA binding protein 1 and 2 (GATA-1 and -2), muscle initiator sequence-19 and -20, and testis specific protein NYD-TSP1. AP-4, GATA-1 and -2 have previously been identified as being associated with common gene targets in OA[10,11]. The down-regulated models share multiple binding sites. These include SP1, GC box, muscle initiator sequence-20, zinc finger protein 161 homolog and AP-2γ. In addition, E2F1, cMyc/Max and AP-2α binding sites are present in two models. Early growth response (EGR) 1 and 4 are only present in model 2. Like the up-regulated model, these models involve transcription factors previously associated with common gene targets in OA. These include Sp1, E2F1, AP-2, and cMyc/Max[12–15]. An interesting finding is the presence of muscle initiator sequence-20 in the three down-regulated models and also found in the up-regulated model. Little is known about this transcription factor including its relevance to OA.

GENES CONTAINING THE PROMOTER MODELS

The presence of binding sites for transcription factors that constitute an OA promoter model in the promoter region of a candidate gene might suggest that the gene is also coordinately regulated with genes involved in OA. We identified novel candidate genes involved in OA by scanning the human genome for the promoter models. We thus identified 723 transcripts containing the up-regulated model. Corresponding numbers for the three down-regulated models were 535, 518 and 799 with 337 transcripts common to all down-regulated models. Of these, 320, 298, 284 and 404 nonredundant locus links were identified for up-regulated model 1 and down-regulated models 1–3, respectively. These sets of transcripts were used to determine if any Gene Ontology biological process was significantly represented. As shown in Table IV, several areas of transcription regulation and development are significant for all transcript sets. Pathway analysis of these sets of transcripts (down-regulated model 2 shown as a representative example) shows a concentration of targets associated with tumor necrosis factor (TNF), insulin and transforming growth factor-β (TGF-β) (Fig. 2). Known associations (regulation, transport, binding, etc.) are represented by links between nodes (gene, transcript or protein product). Seventy of the 284 transcripts identified using down-regulated promoter model 2 (25%) have an association with TGF-β, TNF and/or insulin. Similar results were obtained for other models (data not shown).

CONSERVATION OF PROMOTER MODELS IN DOG, MOUSE AND RAT

Next we investigated whether the binding sites in human promoters are conserved in the corresponding canine and rodent orthologs. We found a higher than expected conservation of promoter models which further strengthens the validity of these promoter models and indicates evolutionary conservation of the processes involved in OA. The canine orthologs of the human genes identified by the promoter

Table III

Conservation of promoter models in upstream regions of dog, mouse and rat. Human genes possessing the promoter models in the 5 kbp upstream were used to identify orthologs in dog, mouse and rat. The number of identified orthologs is shown in column 2. The upstream region of the orthologs was searched for the binding sites constituting the promoter model. The number of orthologs containing the model is shown in column 3. Column 4 shows the expected number of orthologs containing the model as described in Materials and methods and column 5 shows the enrichment as a ratio of observed/expected. Note that all models shown occur significantly in the orthologs of the three species examined suggesting an evolutionary conservation of the gene set involved in OA

Model	Number of genes (dog/mouse/rat)	Observed number of genes containing all PWM hits (dog/mouse/rat)	Expected number of genes containing all PWM hits (dog/mouse/rat)	Ratio of observed/expected for dog/mouse/rat
Up-regulated model 1	165/97/95	102/68/64	9/10/10	11.3/6.8/6.4
Down-regulated model 1	138/84/80	63/52/48	2/5/5	31.5/10.4/9.6
Down-regulated model 2	133/82/75	52/42/39	6/8/8	8.7/5.2/4.8
Down-regulated model 3	181/107/101	43/42/41	4/7/6	10.8/6.0/6.8

Table IV

Gene Ontology biological process table. The P-value indicates whether a given biological process was over represented in the list of selected genes or has been selected by chance. The significant ($P < 0.05$) biological processes represented by human transcripts containing the cis*-binding elements of each model are indicated. Missing values represent those with $P > 0.05$*

Gene Ontology biological process	Promoter model *P*-value			
	Up	Down 1	Down 2	Down 3
Regulation of transcription, DNA-dependent	0.0010	4.3E−06	3.5E−07	1.8E−06
Regulation of transcription	5.5E−07	6.8E−06	5.7E−07	3.1E−06
Transcription, DNA-dependent	0.0017	8.0E−06	6.8E−07	3.8E−06
Transcription	0.0011	1.6E−05	1.4E−06	8.6E−06
Morphogenesis	0.0157	1.1E−04	2.6E−04	0.0052
Nucleobase, nucleoside, nucleotide and nucleic acid metabolism	0.0236	6.0E−04	2.6E−04	4.2E−04
Organogenesis	0.0499	7.5E−04	0.0018	0.0167
Development	0.0112	0.0014	0.0023	0.0279
Central nervous system development	0.0129			
Potassium ion transport	0.0426	0.0201		0.0108
Regulation of transcription from Pol II promoter		0.0371		0.0237
Metabolism			0.0177	
Metal ion transport				0.0433
Neurogenesis			0.0472	

models were identified using the CanFam 1.0 assembly (ftp://ftp/ncbi.nih.gov/genomes/Canis_familiaris/). The following number of canine genes were identified: up-regulated model 1 − 165; down-regulated model 1 − 138; down-regulated model 2 − 133; and down-regulated model 3 − 181. The corresponding numbers of mouse genes identified were 97, 84, 82 and 107 and rat genes were 95, 80, 75 and 101, respectively. We assessed whether the promoter

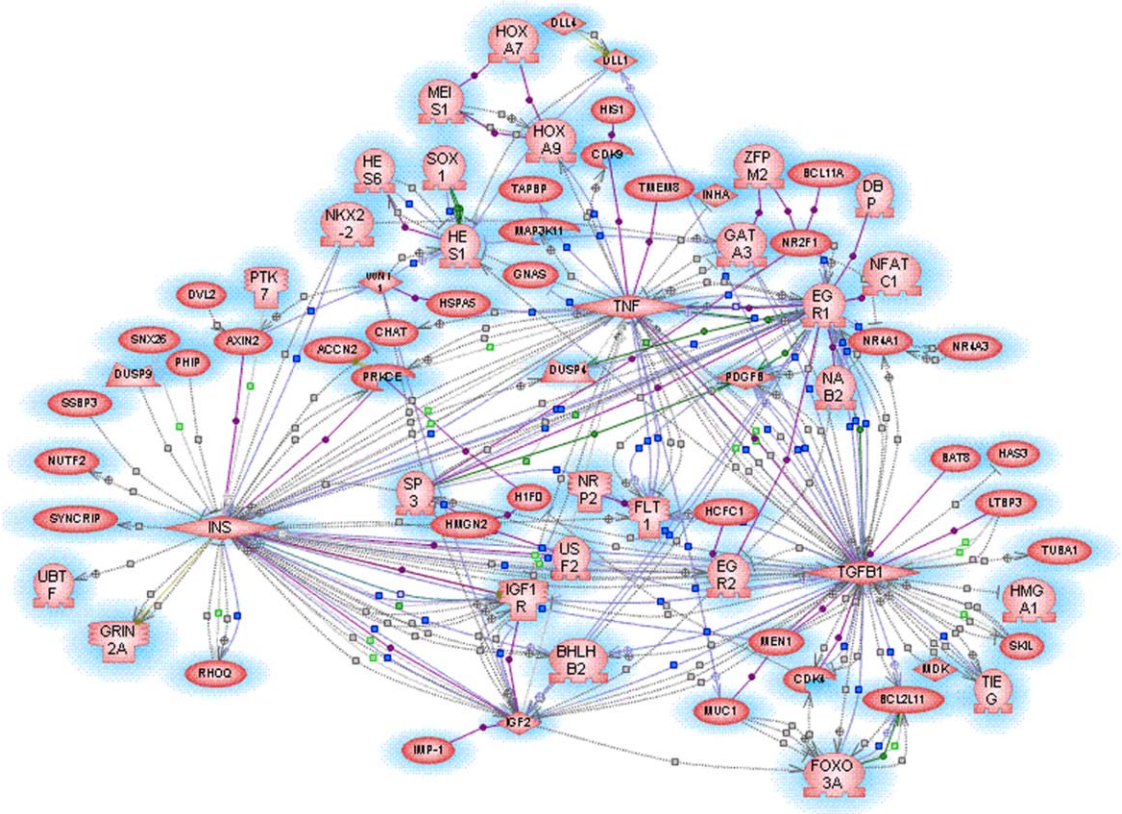

Fig. 2. Human transcripts containing the *cis*-binding elements of down-regulated model 2 associate with TNF, insulin and TGFβ1. Nodes represent transcripts identified and links represent known interactions or modulations from neighboring nodes. Links: regulation, gray square; protein modification, yellow circle; promoter binding, green circle; molecular transport, green open square; molecular synthesis, blue open square; expression, blue square; direct regulation, green square; and binding, purple circle.

regions of the dog, mouse and rat orthologs of the human genes are enriched in the *cis*-elements corresponding to the transcription factors in the corresponding model. As shown in Table III, all of the models are highly enriched in the promoter regions of the transcripts in dog, mouse and rat orthologous to human genes obtained by the direct model searching. For instance, in dog among 165 transcripts, we expect to see only nine transcripts to contain all of the five transcription factor binding sites in their upstream regions, however, we observe 102 such transcripts − 11-fold enrichment. The results for dog, mouse and rat indicate a potentially similar transcriptional control in all three species. In addition, the enrichment is higher for dog vs human than for mouse or rat vs human. This is consistent with a previous analysis of orthologous repeats between human, mouse and dog[16]. Human and dog genome sequences were shown to be more similar to each other than human and mouse.

Discussion

TRANSCRIPTS USED FOR REGULATORY MODEL DEVELOPMENT

The transcripts used for regulatory model development (Table I) were chosen based on their differential expression in canine OA and the inclusion of a set of common binding sites in their upstream regions. These transcripts represent the basis for which the promoter models were developed. Their known function in the disease did not play a role in the model development. Their function, however, may give us a broad understanding of the role these models play in the disease. The transcripts in the up-regulated model include sphingomyelin phosphodiesterase, serine protease 11 (IGF-binding) and chondroadherin, among others. Sphingomyelin phosphodiesterase is better known for its involvement in Niemann−Pick disease but the product from its catalysis of sphingomyelin, ceramide, plays a role as a mediator in apoptosis and matrix degradation in cartilage[17]. Serine protease 11 (IGF-binding) is proposed to cleave IGF-binding proteins, thus regulating the availability of IGFs[18]. IGFs have anabolic effects on chondrocytes. This includes the ability to stimulate the production of proteoglycan[19]. In osteoarthritic cartilage, serine protease 11's RNA and protein expression is increased[20]. The cartilage matrix protein chondroadherin may mediate cell-to-matrix binding. It binds collagen type II and the integrin $\alpha2\beta1$[21]. It may also play a role in the assembly of collagen due to its territorial matrix expression. Taken together, the regulation of these genes may represent a coordinated matrix development, differentiation or reorganization pathway.

Transcripts in the down-regulated models include SRY (sex determining region Y)-box 9 (SOX-9), thrombospondin 1, proprotein convertase subtilisin/kexin type 5 and retinoblastoma binding protein 1 (RBP1), among others. SOX-9 is a transcription factor whose role in cartilage development has been well characterized. SOX-9 plays a major role in chondrogenesis, partly by its ability to induce the expression of chondrocyte-specific genes, such as collagen IIA[22,23]. The expression of SOX-9 is increased in early chondrogenesis but there is a decrease in its expression in metabolically inactive chondrocytes, which associates with a decrease in the proteoglycan matrix and the chondrocyte phenotype[24,25]. In addition, SOX-9 is down-regulated by the E2F1 transcription factor[26]. E2F transcription factors stimulate the expression of genes involved in the cell cycle and possibly tumor suppression. They can also inhibit the expression of genes by forming a complex with the retinoblastoma protein (pRb), which leads to cell cycle arrest (see also RBP1 and E2F1 discussion). The down-regulation of SOX-9 by E2F1 is consistent with our model, since it contains an E2F1 binding site (Table II). Thrombospondin 1 is an extracellular glycoprotein involved in cell−matrix interactions and the activation of TGF-β[27]. In severe OA there is a decrease in thrombospondin associated with a decrease in cartilage matrix[28]. The human thrombospondin gene contains both an AP-2 binding site and a GC box, which is consistent with our findings (Table II)[29]. Proprotein convertase subtilisin/kexin type 5 is a protease that regulates the activity of membrane type-1 MMPs (MT1-MMP)[30]. MT1-MMPs are involved in a latent-form to active-form metalloproteinase cascade leading to the activation of MMP-13 and -9[31]. These MMPs are well known for their cartilage destruction in OA.

RBP1 is a pRb pocket binding protein. It binds pRB/E2F (along with histone deacetylase) in the repression of E2F binding site-related genes[32]. This results in the inhibition of growth promoting protein expression. Although the down-regulation of RBP1 (Table I) and the presence of an E2F1 site in the other down-regulation models (Table II) may tie these models together, the decrease in expression of these genes could be contradictory. Theoretically, the decrease in expression of RBP1 should lead to an increase in expression of E2F binding site-containing genes. This would only apply to genes whose E2F-required expression is inhibited by pRb. SOX-9, as an example, contains an E2F1-binding site and is down-regulated in our model. Its expression, however, is down-regulated by E2F1 in U2OS cells, but not differentially regulated by pRb[26,33]. Future studies will need to be performed to elucidate the interaction of RBP1 and E2F family members in OA.

The transcripts in the down-regulated promoter models are involved in both cartilage matrix degradation and chondrocyte differentiation. While most of these are part of well-characterized biological pathways, these pathways may interact, or may be part of a larger, expanded biological response affecting both matrix turnover and differentiation. These simultaneous biological responses are consistent with the observation that osteoarthritic chondrocytes return to a chondroprogenitor phenotype. Since the genes involved with these responses contain similar *cis*-regulatory elements, their transcriptional regulation would be a mechanism by which these responses could be coordinated.

REGULATORY MODEL TRANSCRIPTION FACTORS

Many of the transcription factors associated with the promoter models identified here have been independently identified in the regulatory regions of known osteoarthritic genes (Table II). As was the case with the differentially expressed transcripts discussed above, the function of the transcription factors and the genes for which they bind did not play a role in model development. However, what is known about their function lends further support for the models representing a coordinately regulated mechanism of matrix development and chondrocyte differentiation.

The transcription factor AP-4, identified in up-regulated model 1, has previously been shown to play a role, along with AP-1, in the interleukin-1 dependent expression of TGF-β[10]. The promoter of the IGF binding protein-2, which binds both IGF-1 and -2, also contains an AP-4 binding site[34]. In addition, the orphan G-protein coupled receptor, CMKLR1, involved in cartilage development, contains

binding sites for AP-4, as well as GATA-1 and -2[11]. TGF-β and the IGFs are potent anabolic factors while the CMKLR1's role beyond its implication in development has not been well defined. These anabolic factors share many of the up-regulated model's *cis*-binding elements. Even though there are few previously known examples of osteoarthritic genes containing these *cis*-binding elements, the function of these genes supports the anabolic side of this model's importance in matrix development, differentiation or reorganization.

The *cis*-binding elements in the down-regulated models are common to many genes involved in OA. As is the case with the transcripts used to develop the models, many of these are involved in chondrocyte differentiation. Binding sites for two AP-2 transcription factors were detected in the down-regulated models. AP-2α was detected in down-regulated models 1 and 3, and AP-2γ was detected in all the three models. AP-2 transcription factors bind their target DNA elements as a homodimer or as heterodimers with other AP-2 family members. AP-2α has been implicated as a negative regulator of chondrocyte differentiation[35]. In chondrocytes, its expression decreases during differentiation while its overexpression leads to a decrease in type II and X collagen, as well as aggrecan. Aggrecan contains both AP-2 and Sp1 sites. Both AP-2 and Sp1 binding to the aggrecan promoter is regulated by TGF-β and p38 MAP kinase in chondrogenic cells[14]. In addition, AP-2 regulates the chondrogenic gene, CD-RAP, which also contains an Sp1 site, in addition to the AP-2 site in its promoter[36]. CD-RAP's expression is also regulated by SOX-9. SOX-9 is one of the transcripts used to develop down-regulated models 1 and 3. The presence of the AP-2 and Sp1 binding sites, along with the differentiation affects of their associated genes, supports the involvement of the down-regulated models in chondrocyte differentiation.

Binding sites for the transcription factor Sp1 are present in all three down-regulated models. Binding sites for Sp1 are fairly common and, as would be expected, many osteoarthritic genes containing an Sp1 site have been identified. These include MMP-9, aggrecan and cartilage-derived retinoic acid-sensitive protein (CD-RAP)[12,37,38]. In addition, Sp1 activity is increased in chondrocytes stimulated with hyaluronan hexasaccharides[39].

Further support of the down-regulated models representing a mechanism of coordinate regulation of differentiation is the presence of E2F1, EGR-1 and cMyc/Max sites. E2F1, as discussed earlier, stimulates the expression of genes involved in the cell cycle and can inhibit the expression of genes by forming a complex with the pRb. The E2F1 binding sites are present in both down-regulated models 1 and 3, which were both developed with the SOX-9 transcript (Tables I and II). The overexpression of E2F1 not only down-regulates SOX-9 but also inhibits chondrocyte differentiation[13,26]. EGR-1 is known for its involvement in cartilage formation[40]. It is also down-regulated in OA[41].

cMyc/Max ties together many of the players discussed thus far via its association with proteins of the anabolic protein TGF-β's signal transduction cascade. The TGF-β signal transduction pathway involves SMAD proteins. cMyc can bind Sp1-SMAD complexes, inhibiting TGF-β induced transcriptional activation of p15(Ink4B)[15]. p15(Ink4B) is a cyclin-dependent kinase 4 (Cdk4) inhibitor. Cdk4 modulates the association of the pRb with E2F transcription factors[42]. In addition, thrombospondin, which is one of the transcripts used in developing the down-regulated models, can activate TGF-β[27]. This series of coordination between cMyc/Max, TGF-β, thrombospondin and Cdk4e can be partially explained by transcriptional regulation due to the presence of the *cis*-binding elements in the down-regulated models. However, further work will be needed to sort out the specific roles they play in chondrocyte differentiation.

GENE ONTOLOGY AND BIOLOGICAL ASSOCIATIONS

As shown in Table IV, human transcripts containing these promoter models are over represented in multiple areas of transcriptional regulation and development. This is indicative of the importance and complexity of transcription in the coordinated regulation of biological pathways, diseases and development, all of which are tied together. The association of transcripts containing these promoter models with development further supports their involvement in this area of OA. This includes chondrocyte or matrix development, differentiation and reorganization.

TNF, TGF-β, insulin, and their associated family members are key players in OA as well as chondrocyte differentiation and matrix reorganization. The involvement of the models developed here with these key players is supported by the magnitude of associations of transcripts driven by these models and these three key players (Fig. 2). It is clear from this broad view, that the models developed here involve these players. This may not be direct regulation (i.e., *cis*-binding), but may be down-stream events through signal transduction cascades or other yet to be characterized biological pathways.

We have sought out to identify common gene regulatory *cis*-elements associated with subsets of differentially expressed osteoarthritic transcripts. Four models have been developed and are conserved in human, dog, mouse and rat. The conservation gives further validity of the models, since spurious matches would be less likely to be conserved in the different species. Additionally, conservation may also indicate a similarity in disease mechanisms among the different species, further establishing their use as model organisms. Taking into account the transcripts used, the transcription factors associated with the *cis*-binding elements and other transcripts driven by these models, they may represent a coordinated regulation mechanism of chondrocyte differentiation or extracellular matrix reorganization. The subtle differences between the three down-regulated models may provide an efficient means for transcriptional regulation under varying conditions, which still remains to be explored.

References

1. Pfizer Animal Health. Proprietary market research. Veterinary sample size: 200. Available from: http://www.rimadyl.com; 1996.
2. Lawrence RC, Hochberg MC, Kelsey JL, McDuffie FC, Medsger TA, Felts WR, *et al*. Estimates of the prevalence of selected arthritic and musculoskeletal diseases in the United States. J Rheumatol 1989;16:427–41.
3. Middleton RP, Hannah SS, Perroud BN, McDonald KK. Gene expression profiling of osteoarthritis from canine chondrocytes. The development of a canine OA microarray chip. Trans Orthop Res Soc 2003;28:741.
4. Hannenhalli S, Levy S. Predicting transcription factor synergism. Nucleic Acids Res 2002;30:4278–84.
5. Hannenhalli S, Levy S. Transcriptional regulation of protein complexes and biological pathways. Mamm Genome 2003;14:611–9.

6. Wingender E, Dietze P, Karas H, Knuppel R. TRANSFAC: a database on transcription factors and their DNA binding sites. Nucleic Acids Res 1996;24:238–41.
7. Levy S, Hannenhalli S. Identification of transcription factor binding sites in the human genome sequence. Mamm Genome 2002;13:5104.
8. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990; 215:403–10.
9. Mutch DM, Simmering R, Donnicola D, Fotopoulos G, Holzwarth JA, Williamson, *et al*. Impact of commensal microbiota on murine gastrointestinal tract gene ontologies. Physiol Genomics 2004;19:22–31.
10. Andriamanalijaona R, Felisaz N, Kim SJ, King-Jones K, Lehmann M, Pujol JP, *et al*. Mediation of interleukin-1 beta-induced transforming growth factor beta1 expression by activator protein 4 transcription factor in primary cultures of bovine articular chondrocytes: possible cooperation with activator protein-1. Arthritis Rheum 2003;48(6):1569–81.
11. Martensson UE, Owman C, Olde B. Genomic organization and promoter analysis of the gene encoding the mouse chemoattractant-like receptor, CMKLR1. Gene 2004;328:167–76.
12. Bosserhoff AK, Kondo S, Moser M, Dietz UH, Copeland NG, Gilbert DJ, *et al*. Mouse CD-RAP/MIA gene: structure, chromosomal localization, and expression in cartilage and chondrosarcoma. Dev Dyn 1997;208:516–25.
13. Scheijen B, Bronk M, van der Meer T, Bernards R. Constitutive E2F1 overexpression delays endochondral bone formation by inhibiting chondrocyte differentiation. Mol Cell Biol 2003;23:3656–68.
14. Tuli R, Seghatoleslami MR, Tuli S, Howard MS, Danielson KG, Tuan RS. p38 MAP kinase regulation of AP-2 binding in TGF-beta1-stimulated chondrogenesis of human trabecular bone-derived cells. Ann N Y Acad Sci 2002;961:172–7.
15. Feng XH, Liang YY, Liang M, Zhai W, Lin X. Direct interaction of cMyc with Smad2 and Smad3 to inhibit TGF-beta-mediated induction of the CDK inhibitor p15(Ink4B). Mol Cell 2002;9:133–43.
16. Kirkness EF, Bafna V, Halpern AL, Levy S, Remington K, Rusch DB, *et al*. The dog genome: survey sequencing and comparative analysis. Science 2003;301:1898–903.
17. Sabatini M, Rolland G, Leonce S, Thomas M, Lesur C, Perez V, *et al*. Effects of ceramide on apoptosis, proteoglycan degradation, and matrix metalloproteinase expression in rabbit articular cartilage. Biochem Biophys Res Commun 2000;267:438–44.
18. Zumbrunn J, Trueb B. Primary structure of a putative serine protease specific for IGF-binding proteins. FEBS Lett 1996;398:187–92.
19. Mi Z, Ghivizzani SC, Lechman ER, Jaffurs D, Glorioso JC, Evans CH, *et al*. Adenovirus-mediated gene transfer of insulin-like growth factor 1 stimulates proteoglycan synthesis in rabbit joints. Arthritis Rheum 2000;43:2563–70.
20. Hu S, Carozza M, Klein M, Nantermet P, Luk D, Crowl RM. Human HtrA, an evolutionarily conserved serine protease identified as a differentially expressed gene product in osteoarthritic cartilage. J Biol Chem 1998;273:34406–12.
21. Månsson B, Wenglén C, Mörgelin M, Saxne T. Association of chondroadherin with collagen type II. J Biol Chem 2001;276:32883–8.
22. Zhao Q, Eberspaecher H, Lefebvre V, De Crombrugghe B. Parallel expression of Sox9 and Col2a1 in cells undergoing chondrogenesis. Dev Dyn 1997;209:377–86.
23. Lefebvre V, Huang W, Harley VR, Goodfellow PN, de Crombrugghe B. SOX9 is a potent activator of the chondrocyte-specific enhancer of the Proα1(II) collagen gene. Mol Cell Biol 1997;17:2336–46.
24. Salminen H, Vuorio E, Saamanen AM. Expression of Sox9 and type IIA procollagen during attempted repair of articular cartilage damage in a transgenic mouse model of osteoarthritis. Arthritis Rheum 2001;44:947–55.
25. Uusitalo H, Salminen H, Vuorio E. Activation of chondrogenesis in response to injury in normal and transgenic mice with cartilage collagen mutations. Osteoarthritis Cartilage 2001;9:S174–9.
26. Müller H, Bracken AP, Vernell R, Moroni MC, Christians F, Grassilli E, *et al*. E2Fs regulate the expression of genes involved in differentiation, development, proliferation, and apoptosis. Genes Dev 2001;15:267–85.
27. Schultz-Cherry S, Murphy-Ullrich JE. Thrombospondin causes activation of latent transforming growth factor-β secreted by endothelial cells by a novel mechanism. J Cell Biol 1993;122:923–32.
28. Pfander D, Cramer T, Deuerling D, Weseloh G, Swoboda B. Expression of thrombospondin-1 and its receptor CD36 in human osteoarthritic cartilage. Ann Rheum Dis 2000;59:448–54.
29. Laherty CD, Gierman TM, Dixit VM. Characterization of the promoter region of the human thrombospondin gene. DNA sequences within the first introns increase transcription. J Biol Chem 1989;264:11222–7.
30. Yana I, Weiss SJ. Regulation of membrane type-1 matrix metalloproteinase activation by proprotein convertases. Mol Biol Cell 2000;11:2387–401.
31. Dreier R, Grassel S, Fuchs S, Schaumburger J, Bruckner P. Pro-MMP-9 is a specific macrophage product and is activated by osteoarthritic chondrocytes via MMP-3 or a MT1-MMP/MMP-13 cascade. Exp Cell Res 2004;297:303–12.
32. Lai A, Kennedy BK, Barbie DA, Bertos NR, Yang XJ, Theberge M, *et al*. RBP1 recruits the mSIN3-histone deacetylase complex to the pocket of retinoblastoma tumor suppressor family proteins found in limited discrete regions of the nucleus at growth arrest. Mol Cell Biol 2001;21:2918–32.
33. Vernell R, Helin K, Müller H. Identification of target genes of the $p16^{INK4A}$-pRB-E2F pathway. J Biol Chem 2003;278:46124–37.
34. Badinga L, Song S, Simmen RC, Simmen FA. A distal regulatory region of the insulin-like growth factor binding protein-2 (IGFBP-2) gene interacts with the basic helix-loop-helix transcription factor, AP-4. Endocrine 1998;8:281–9.
35. Huang Z, Xu H, Sandell L. Negative regulation of chondrocyte differentiation by transcription factor AP-2alpha. J Bone Miner Res 2004;19:245–55.
36. Xie WF, Kondo S, Sandell LJ. Regulation of the mouse cartilage-derived retinoic acid-sensitive protein gene by the transcription factor AP-2. J Biol Chem 1998; 273:5026–32.
37. Sato H, Seiki M. Regulatory mechanism of 92 kDa type IV collagenase gene expression which is associated with invasiveness of tumor cells. Oncogene 1993;8:395–405.
38. Watanabe H, Gao L, Sugiyama S, Doege K, Kimata K, Yamada Y. Mouse aggrecan, a large cartilage

proteoglycan: protein sequence, gene structure and promoter sequence. Biochem J 1995;308:433–40.
39. Ohno S, Im HJ, Knudson CB, Knudson W. Hyaluronan oligosaccharide-induced activation of transcription factors in bovine articular chondrocytes. Arthritis Rheum 2005;52:800–9.
40. Kamochi H. cDNA macroarray analysis of genes responsible for cartilage formation. Kokubyo Gakkai Zasshi 2004;71:27–37.
41. Wang F, Connor JR, Dodds RA, James IE, Kumar S, Zou C, *et al*. Differential expression of Egr-1 in osteoarthritic compared to normal adult human articular cartilage. Osteoarthritis Cartilage 2000;8:161–9.
42. Ezhevsky SA, Nagahara H, Vocero-Akbani AM, Gius DR, Wei MC, Dowdy SF. Hypo-phosphorylation of the retinoblastoma protein (pRb) by cyclin D: Cdk4/6 complexes results in active pRb. Proc Natl Acad Sci U S A 1997;94:10699–704.